WILEY SERIES IN PROBABILITY AND STATISTICS
TEXTS AND REFERENCES SECTION

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *Vic Barnett, Ralph A. Bradley, Noel A. C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J. B. Kadane, David G. Kendall, David W. Scott, Bernard W. Silverman, Adrian F. M. Smith, Jozef L. Teugels, Geoffrey S. Watson; J. Stuart Hunter, Emeritus*

A complete list of the titles in this series appears at the end of this volume.

Applied Regression Analysis

THIRD EDITION

Norman R. Draper

Harry Smith

Library of Congress Cataloging-in-Publication Data:

Draper, Norman Richard.
Applied regression analysis / N.R. Draper, H. Smith. — 3rd ed.
p. cm. — (Wiley series in probability and statistics. Texts and references section)
"A Wiley-Interscience publication."
Includes bibliographical references (p. –) and index.
ISBN 0-471-17082-8 (acid-free paper)
1. Regression analysis. I. Smith, Harry, 1923– . II. Title. III. Series.
QA278.2.D7 1998

Contents

Preface to the Third Edition

The second edition had 10 chapters; this edition has 26. On the whole (but not entirely) we have chosen to use smaller chapters, and so distinguish more between different types of material. The tabulation below shows the major relationships between second edition and third edition sections and chapters.

Material dropped consists mainly of second edition Sections 6.8 to 6.13 and 6.15, Sections 7.1 to 7.6, and Chapter 8. New to this edition are Chapters 16 on multicollinearity, 18 on generalized linear models, 19 on mixture ingredients, 20 and 21 on the geometry of least squares, 25 on robust regression, and 26 on resampling procedures. Small revisions have been made even in sections where the text is basically unchanged. Less prominence has been given to printouts, which nowadays can easily be generated due to the excellent software available, and to references and bibliography, which are now freely available (either in book or computer form) via the annual updates in *Current Index to Statistics*. References are mostly given in brief either in situ or close by, at the end of a section or chapter. Full references are in a bibliography but some references are also given in full in sections or within the text or in exercises, whenever this was felt to be the appropriate thing to do. There is no precise rule for doing this, merely the authors' predilection. Exercises have been grouped as seemed appropriate. They are intended as an expansion to the text and so most exercises have full or partial solutions; there are a very few exceptions. One hundred and one true/false questions have also been provided; all of these are in "true" form to prevent readers remembering erroneous material. Instructors can reword them to create "false" questions easily enough. Sections 24.5 and 24.6 have some duplication with work in Chapter 20, but we decided not to eliminate this because the sections contain some differences and have different emphases. Other smaller duplications occur; in general, we feel that duplication is a good feature, and so we do not avoid it.

Our viewpoint in putting this book together is that it is desirable for students of regression to work through the straight line fit case using a pocket calculator and then to proceed quickly to analyzing larger models on the computer. We are aware that